Sriram Mohan

Comp. Sci. & Software Engineering
Rose-Hulman Institute of Tech.
5500 Wabash Ave.
Terre Haute IN 47803

Phone: 812.877.8819
Fax: 812.877.6260
mohan@rose-hulman.edu
http://www.rose-hulman.edu/~mohan/

Education

Ph.D. in Computer Science, May 2007.
Indiana University, Bloomington, USA.
*D*issertation: "ACXESS - A Framework for Designing Specifying and Enforcing Access Control for XML".
Minor - Human Computer Interaction.

M.S in Computer Science, May 2003.
Indiana University, Bloomington, USA.

B.E. in Computer Science and Engineering, May 2001.
University of Madras, Madras, India.
*T*hesis:"NetRep- Internet Call Center for a Distributed Environment".

Research Experience

XML ACCESS CONTROL

I am a key member of the ACXESS [1,2,3](Access Control for XML with Enhanced Security Specifications) project, in which we are building a framework for designing, specifying and enforcing access control for XML. My research has focused on the design and development of a constraint specification language for XML.

Being able to express and enforce role-based access control on XML data is a critical component of XML data management. However, given the semi-structured nature of XML, this is non-trivial, as access control can be applied on the values of nodes as well as on the structural relationship between nodes. ACXESS introduces a notion of security views to enforce access control for XML data repositories. I have introduced an algebraic security view specification language - SSX[1] for DBAs to specify access constraints for different user groups. ACXESS does not materialize the security views and enforces the access constraints via query rewrites.

To facilitate query answering and rewriting, ACXESS introduces an internal representation - Security Annotated Schema (SAS) in the form of annotations to represent the schema transformation specified by a SSX sequence. Annotations are associated with the element node that was modified and reflects the actual changes performed on the original schema tree structure.

Security constraints are enforced by rewriting user queries on the security view to a secure XQuery expression. I have introduced Secure Query Rewrite (SQR)[3] - a set of rules to rewrite a user XPath query on the security view into an equivalent XQuery expression against the original data, with the guarantee that the users only see information in the view but not any data that was blocked.

Access control languages have to worry about inference as information can still be inferred by running a set of queries, over a period of time, to enable user groups with lower security level to correctly identify data that is visible to user groups with higher security level. The presence of access constraints on structure makes anti-inference non-trivial. I am currently working on developing anti-inference algorithms to minimize information leakage. Possible solutions include breaking document order, detecting functional dependencies among sibling nodes and monitoring user queries to prevent information leakage.

Conceptual Modeling for XML

Many organizations are starting to store data in XML and using XML as an intermediate format for publication and use of these documents. Most database systems have support for storing data in XML or internally representing XML data for storage. However, XML does not have a suitable mechanism for intuitively creating a conceptual model for the data and cannot automatically or semi-automatically generate the schema for the actual data storage [4]. XML features such as complex elements and heterogeneity complicate the development of a conceptual model for XML.

In this context, I have introduced XER[7,8](Extensible Entity- Relationship) model, a conceptual modeling approach that can describe XML document structures in a simple visual form reminiscent of the ER model, and has the capability to automatically generate XML document type definitions and schema from such structures.

The XER entity is the basic conceptual object in XER [7] and represents complex XML concepts. XER attributes are properties of entities that are usually atomic but can also be optional or multi-valued. Attributes are shown in the model by placing the names of the attributes in the body of the entity. Relationships, which denote a connection between two or more entities, are introduced in XER when a complex entity contains a complex element as one of its sub-elements. XER supports ER concepts such as weak entities, ternary relationships, aggregations and generalizations. The XER diagram [7] also supports more or less every facet available in XML Schema. XER has been prototyped [8] using XSLT and Dia and is capable of creating new XER diagrams (using a specially developed stencil), performing up-translations - conversion of XML schema to XER diagrams and down-translations - converting XER back to XML schema.

A lab study was conducted to empirically evaluate XER with respect to currently existing methodologies to model XML. Statistical analysis demonstrate XER to be superior to other modeling artifacts for modeling XML structures, significantly improving accuracy, efficiency and user satisfaction.

DBMS as an IR Engine

Can a system designed primarily for the purpose of database-type storage and retrieval be used for information-retrieval tasks? DocBase, a prototype database system developed initially for SGML, and adapted to work with XML, was used for the

purpose of answering the queries. DocBase uses DSQL, an adaptation of SQL to provide a mechanism for querying XML using existing database and indexing technologies. DocBase was benchmarked as a part of the INEX initiative. Although the benchmark [5] revealed the limitations of database query languages for classic information retrieval tasks, it also demonstrated that several interesting results can be obtained by using database query languages for information retrieval, especially for queries involving both content and structure. The Benchmark [5] results demonstrate the adaptability and scalability of a database system for processing IR queries.

XML Database Management System

The emergence of XML as a predominant format for document representation has necessitated the requirement of effective storage and retrieval of XML documents. DocBase is a complete solution architecture for XML, including a collection of several novel methods related to XML. The architecture, and the implemented system, *DocBase*, uses a flexible storage and indexing technique to allow highly expressive queries without the necessity of mapping documents to other database formats. DocBase introduces a SQL like language - DSQL for querying XML databases. My contributions include the implementation of the 'Project' and the 'Join' operator and an extensive performance benchmarking of DocBase using several widely available XML benchmarks.

Pervasive Audience Participation System

In recent years, the Olympic games have undergone scrutiny for problems due to subjectivity in judging events such as ice skating. There is a need for systems that can highlight this discrepancy and help sustain audience involvement and interest with the Olympics. The audience participate in the Olympics by rewarding athletic performances with cheers and applause. While this auditory feedback is appreciated by the athletes, it is hard to quantify the response into a meaningful or quantifiable score. My research introduced 'WIN', a pervasive audience participation system [6] which senses natural forms of audience appreciation (clapping and waving), and utilizes this as a rough but instinctive measure of their response to the athletes. By showing the difference between the audience and the judges' scores, it is expected that the scandals currently plaguing the games can be highlighted. 'WIN' moves beyond engaging just the individual, it views the audience as a collective unit and enhances social interactions by involving them in entertaining activities and games. 'WIN' attempts to become a part of the audience, without distracting them,disappearing into their periphery; coming back to their attention only to connect them with their emotions, with their participation and with other people.

Risk Management

The most important thing for a project is to focus on critical success factors. Due to the influence of previous document driven management guidelines, projects focus on activities that are not critical towards success. One of the most important contributions of software risk management is to create this focus on critical success factors and to provide techniques and a framework that enables the project to deal

with them. However risk management is not a cookbook approach. To handle all the complex people oriented and technology driven success factors in projects a great measure of human judgment is required. Various risk assessment and risk control strategies were surveyed to determine a foundation layer of capabilities that have to be ingrained into a software engineer to implement the risk-oriented approach. Based on the survey my research introduced methodologies to introduce risk management as a part of the software engineering curriculum.

Refereed Publications

1.*A Generalized Rewrite based Framework for Access Control for XML* , with Arijit Sengupta and Yuqing Wu, atnternational Conference on Knowledge-Based and Intelligent Information & Engineering Systems. 2007

2.*IPAC - An Interactive Approach to Access Control for Semi-Structured Data* with Yuqing Wu. Demonstration at the International Conference on Very Large Databases 2006.

3.*ACXESS - Access Control for XML with Enhanced Security Specifications*, with Arijit Sengupta, Yuqing Wu and Jonathan Klinginsmith, Demonstration at the International Conference on Data Engineering, 2006.

4.*Access Control for XML - A Dynamic Query Rewriting Approach*, with Arijit Sengupta and Yuqing Wu, at the ACM Conference on Information and Knowledge Management, 2005.

5. *Conceptual Modeling for XML - A Myth or Reality*, with Arijit Sengupta. Database Modeling for Industrial Data Management: Emerging Technologies and Applications. Idea Group Inc. 2005.

6.*DocBase - The INEX Evaluation Experience*, with Arijit Sengupta. Initiative for the Evaluation of XML Retrieval 2004, Springer LNCS.

7.*WeInteract - A Pervasive Audience Participation System*, with Vasudha Chandrasekaran, Sidharth Saxena and Om Prakash Pathipaka. ACM Special Interest Group on Computer Human Interaction 2004.

8.*Extensible Entity Relationship Modeling,*with Arijit Sengupta and Rahul Doshi, in J. Harnad et al. Eds. Proceedings of the XML 2003 Conference.

9.*Conceptual Modeling for XML using XER*, with Arijit Sengupta. Workshop on Information Technology and Systems (WITS) 2003.

Under Review

A Generalized Access Control Framework for XML Applications with Yuqing Wu. Submitted.

A Framework for Access Control for XML - Beyond Subtree Hiding, with Arijit Sengupta and Yuqing Wu. Submitted.

XER - Extensible Entity Relationship Modeling, with Arijit Sengupta. Submitted.

DocBase - A Document Centric Approach to Managing XML Data, with Arijit Sengupta and Ramesh Venkatraman. Submitted.

Posters

ACXESS - Access Control for XML with Enhanced Security Specifications, with Arijit Sengupta, Yuqing Wu. Second Midwest Database Research Symposium, Chicago, Illinois, 2005.

Data Modeling and Optimizing XML Databases, with Arijit Sengupta. First Midwest Database Research Symposium, Chicago, Illinois, 2004.

White Papers

Access Control for XML - A Dynamic Query Rewriting Approach, with Arijit Sengupta and Yuqing Wu . Technical Report No. 609, Computer Science, Indiana University.2005.

Formal and conceptual models for XML structures - the past, present and future with Arijit Sengupta. Technical Report no. 137-1, Information Series Department Working Paper Series. Indiana University, April 2003.

Risk Management, An Annotated Bibliography, Masters Thesis, Computer Science, Indiana University.2003.

NetRep - Internet Call Center for a Distributed Environment, Undergraduate Thesis, Computer Science and Engineering, University of Madras.2001.